



## An Introduction to Bioinformatics Resources and their Practical Applications

Medha Bhagwat, PhD  
Coordinator

NIH Library Bioinformatics Support Program  
bhagwat@mail.nih.gov



## *(Really)* An Introduction to *(Few)* Bioinformatics Resources and their Practical Applications

Medha Bhagwat, PhD  
Coordinator

NIH Library Bioinformatics Support Program  
bhagwat@mail.nih.gov

## Bioinformatics

### Variety of definitions

Luscombe et al. *Method Inform Med* 2001; 40:346-58

Bioinformatics is conceptualizing **biology in terms of molecules**  
(in the sense of Physical chemistry)  
and applying "**informatics techniques**"  
(derived from disciplines such as applied  
math, computer science and statistics)  
to **understand and organize** the information associated  
with these molecules, **on a large scale**.

Bioinformatics is a management information system for  
molecular biology and has many practical applications.

## Bioinformatics

### I. Organize data in databases

- access current data
- submit new data

### II. Develop tools and resources to analyze data

### III. Interpret data in a biologically useful manner

- global analysis of data to uncover  
common principles that apply across  
many systems



# Bioinformatics

next gen

genomics

proteomics

structure

pathways

?

What does the NIH community need?

NIH LIBRARY

## Identifying and Supporting NIH Researchers' Bioinformatics Needs

October 6  
11AM-1PM

RSCHSUPP – 4

NIH LIBRARY

NIH Library  
Amazing Research. Amazing Help.

For the Public | FAQ | Help | Site Map

Home | Library Services | Research Tools | Resource Training | About Us

## Research Tools

NIH Library | Research Tools | Bioinformatics

**Bioinformatics**

Ask A Librarian

Search Library Site

**Quick Links**

- Bioinformatics
- Going Green
- Online Books
- Online Catalog
- Online Databases
- Online Journals
- Order a Document
- PubMed @ NIH
- Remote Access
- Scopus
- Web of Science
- Sign-Up For
- Library Email News
- PubMed Direct
- Research Updates

**Announcements**

- NIH Library Classes: August and September 2010
- Launch of Beta Search Engine: First Step
- Locked Carnis 2010
- NIH Library All Announcements
- [View All Announcements](#)

**Upcoming Training Classes**

- PubMed: Understanding the Basics Oct 13
- EndNote X3: Managing Your Search Results Oct 13
- EndNote X3: Managing Your Search Results Nov 19

[Request a Tutorial](#)

Dr. Yi Ding, a Clinical Center postdoc, and Dr. Medha Bhagwat, the NIH Library Bioinformatics Support Program Coordinator, discuss how a library tutorial with Medha saved Ding time in being able to successfully clone a gene promoter sequence.

The NIH Library's Bioinformatics Support Program was developed to provide researchers with powerful tools to analyze and understand the biologic significance of a variety of data. The program is conducted by an expert bioinformatics trainer, Medha Bhagwat, and consists of the following:

- One-on-one consultation
- Classroom training
- Analysis Tools and Databases
- Online Tutorials
- Bioinformatics Program Staff
- Testimonials

"Bhagwat is especially helpful to researchers, because she was a bench scientist and she knows what scientists need. She knows computers, writes scripts, and extracts data from databases."

Hary Ann Robinson, Microbiologist  
National Institute of Allergy and Infectious Diseases (NIAID)

<http://nihlibrary.nih.gov/bioinformatics>

## Outline

- Databases and tools (free)
- Licensed resources (from the NIH Library)
- Specific examples
- Training and additional help

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

# Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > Database Summary Paper Categories

## 2010 NAR Database Summary Paper Category List

- Nucleotide Sequence Databases
- RNA sequence databases
- Protein sequence databases
- Structure Databases
- Genomics Databases (non-vertebrate)
- Metabolic and Signaling Pathways
- Human and other Vertebrate Genomes
- Human Genes and Diseases
- Microarray Data and other Gene Expression Databases
- Proteomics Resources
- Other Molecular Biology Databases
- Organelle databases
- Plant databases
- Immunological databases

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

<http://www.oxfordjournals.org/nar/database/c/>

LIBRARY  
NIH

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

# Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > Web Server Summary Categories

## NAR Web Server Categories List

- Computer Related
- DNA
- Education
- Expression
- Human Genome
- Literature
- Model Organisms
- Other Molecules
- Protein
- RNA
- Sequence Comparison

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Summary List](#)
- ▶ [Search Summaries](#)
- ▶ [Bioinformatics Links Directory](#)

[http://www.oxfordjournals.org/nar/webserver/c](http://www.oxfordjournals.org/nar/webserver/c/)

# Nucleic Acids Res

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSC

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > W

NAR Web Server C

Computer Related  
DNA

- Annotations
- Gene Prediction
- Mapping and Assem
- CAPS Sequences
- eBioinformatics
- EGassembler
- ESTpass
- FPC
- Genome Centre
- Genome Databa
- ICE - internet C
- MagicViewer
- MGIP
- NCBI Clone Res
- Phred/Phrap/Gc
- Projector 2
- PromoSer
- Staden Package
- TIGR Software
- TmPrime
- TOM
- WebPrInSeS
- ZOOM Lite

# Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > [Web Server Summary 3338](#)

◀ PREVIOUS    NEXT ▶

## ZOOM Lite

NAR Web Server Collection entry number 3338

<http://bioinform.com/zoom/lite/>

Contact [binma@uwaterloo.ca](mailto:binma@uwaterloo.ca)

### Description

ZOOM Lite is a web tool for efficient read mapping and result visualization for next generation sequencing data. Capable of handling data from Illumina and AB SOLID platforms, and single- and paired-end reads.

### Classification

- Category: DNA
- Subcategory: Mapping and Assembly
- Category: Expression
- Subcategory: Transcript Expression & Microarrays

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Summary List](#)
- ▶ [Search Summaries](#)
- ▶ [Bioinformatics Links Directory](#)

Special Issues: Current progress in Bioinformatics 2010

Volume 11 | Number 1 | Jan 2010


**Briefings in Bioinformatics**


www.bib.oxfordjournals.org

## Special Issues

## Current progress in Bioinformatics 2010

## Second Generation Sequencing

Computational Molecular Biology at NIH	
<a href="#">Index</a> <a href="#">Sequence Analysis</a> <a href="#">Structure Analysis</a> <a href="#">Systems Biology</a> <a href="#">Documentation</a> <a href="#">Databases</a> <a href="#">Reference</a> <a href="#">Software</a> <a href="#">Other</a>	
<p><b>Main Index</b></p> <p><a href="#">Main Index</a></p> <p><a href="#">Sequence Analysis</a> - tools</p> <p><a href="#">Structure Analysis</a> - specific</p> <p><a href="#">Systems Biology</a> - for large</p> <p><a href="#">Documentation and Inform</a></p> <p><a href="#">Databases</a> - for molecular b</p> <p><a href="#">Molecular Biology Desk R</a></p> <p><a href="#">Software Repositories</a> - ac</p> <p><a href="#">Other Molecular Biology R</a></p>	<p><b>Sequence Analysis</b></p> <hr/> <p><b>Multifunctional Tools</b></p> <p><a href="#">EMBOSS</a> - A comprehensive set of sequence analysis programs that handles all sequence formats (NIH only)</p> <p><a href="#">NIH EMBOSS-lite</a> - interface to the more popular EMBOSS utilities (NIH only)</p> <p><a href="#">NIAID BioCluster: Bioinformatics Portal</a> - Access to 200 bioinformatics applications and custom workflows using a convenient web-based interface (NIH only)</p> <p><a href="#">ABS</a> - Analytical Biostatistics Section (NIH only)</p> <p><a href="#">DAVID</a> - Database for Annotation, Visualization and Integrated Discovery (NIH only)</p> <p><a href="#">EBI Services</a> - EMBL European Bioinformatics Institute Services</p> <p><a href="#">EMBL Heidelberg Tools</a> - EMBL Heidelberg</p> <p><a href="#">Institute for Genomics and Bioinformatics</a> - Structural Proteomics</p> <p><a href="#">Galaxy</a> - Genomic metaserver with access to a multitude of databases, at Penn State Univ.</p> <p><a href="#">RPBS</a> - Ressource Parisienne en Bioinformatique Structurale</p> <p><b>Text-Based Query of Sequence Databases</b></p> <p><a href="#">Entrez Cross Database Query</a> - one-stop shopping for NCBI databases</p> <p><a href="#">Sequence Retrieval System</a> - at EBI (UniProt)</p> <p><a href="#">SwissProt Search</a> - at FxPASy.org (UniProt)</p> <p><a href="#">PIR Searches</a> - at Georgetown University (UniProt)</p> <p><a href="#">Molecules To Go</a> - text-based interface to the PDB on Helix Systems</p> <p><a href="#">PDB Searches</a> - advanced search at the PDB</p> <p><a href="#">Nucleic Acid Database Search</a> - at Rutgers</p> <p><a href="#">Annotation-Modules</a> - tool for finding significant combinations of multisource annotations, at CIC bioGUNE</p> <p><a href="#">PDBsum</a> - At-a-glance overview of PDB structures, at EBI</p> <p><b>Sequence Format Conversion</b></p> <p><a href="#">Format Conversion via EMBOSS seqret</a> - at Helix Systems (NIH only)</p> <p><a href="#">READSEQ Sequence Conversion</a> - at CIT/NIH</p> <p><b>Database Similarity/Homology Searches</b></p> <p><a href="#">Basic Local Alignment Search Tool (BLAST)</a> - the gold standard for homology searches</p>
	

Computational Molecular Biology at NIH	
<a href="#">Index</a> <a href="#">Sequence Analysis</a> <a href="#">Structure Analysis</a> <a href="#">Systems Biology</a> <a href="#">Documentation</a> <a href="#">Databases</a> <a href="#">Reference</a> <a href="#">Software</a> <a href="#">Other</a>	
<p><b>Main Index</b></p> <p><a href="#">Main Index</a></p> <p><a href="#">Sequence Analysis</a> - tools at NIH and around the world</p> <p><a href="#">Structure Analysis</a> - specifically for structural biologists</p> <p><a href="#">Systems Biology</a> - for large-scale interaction networks</p> <p><a href="#">Documentation and Information</a> - about tools and programs</p> <p><a href="#">Databases</a> - for molecular biology</p> <p><a href="#">Molecular Biology Desk Reference</a> - collection of basic information</p> <p><a href="#">Software Repositories</a> - access to molecular biology and sequence analysis software</p> <p><a href="#">Other Molecular Biology Resources</a> - major websites for molecular biology</p>	
<p>Searchable and an ability to add feedback</p>	
	

## Some Free Resources for Next Gen Analysis

The screenshot shows the NIH Library Research Tools page. The main heading is "Free External Resources for Next Generation Sequencing". The page is organized into several sections:

- Alignment:**
  - BFAST - Mapping to reference sequences
  - Bowtie - Aligner using Burrows-Wheeler algorithm
  - BWA - Burrows-Wheeler Aligner
  - MAQ - Mapping and Assembly with Qualities
- Assembly:**
  - Abyss - Assembly by Short Sequences (de novo)
  - AMOSmp-shortreads - Assemble reads from an organism by mapping to reference from another organism
  - MIRA - whole genome shotgun and EST sequence assembler
  - mosaik-assembler - Reference-guided assembler
  - Velvet - Sequence assembler for very short reads
- ChIP-Seq:**
  - MACS - Model-based Analysis for ChIP-Seq
  - FindPeaks - Analysis of ChIP experiments
- RNA-Seq:**
  - Cufflinks - Transcript assembly, gene expression, regulation
  - Myrna - Cloud computing solution for differential gene expression in large RNA-Seq datasets
- Coding Sequence Prediction:**
  - G-MoR-Se - Gene Modeling using RNA-Seq
  - TopHat - Splice junction mapper for RNA-Seq
  - Multi-use

On the right side, there are sections for "Announcements" and "Upcoming Training Classes". A "Quick Links" sidebar is visible on the left.

## Some Free Resources for Next Gen Analysis

<http://bowtie-bio.sourceforge.net/myrna/index.shtml>

**Myrna**  
Cloud-scale differential gene expression for RNA-seq

JOHNS HOPKINS BLOOMBERG SCHOOL OF PUBLIC HEALTH

OSI certified

Myrna is a cloud computing tool for calculating differential gene expression in large RNA-seq datasets. Myrna uses Bowtie for short read alignment and R/Bioconductor for interval calculations, normalization, and statistical testing. These tools are combined in an automatic, parallel pipeline that runs in the cloud (Elastic MapReduce in this case) on a local Hadoop cluster, or on a single computer, exploiting multiple computers and CPUs wherever possible.

**Related Tools**


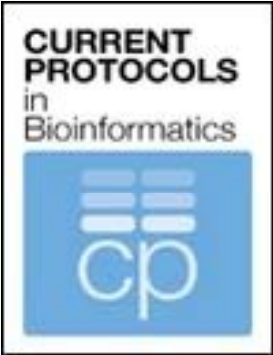
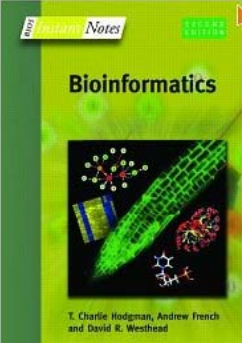
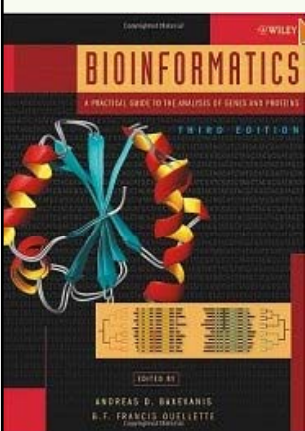
- Bowtie:** Ultrafast short read alignment
- Crossbow:** Cloud-scale genotyping
- Contrail:** Cloud-based *de novo* assembly
- Hadoop:** Open Source MapReduce
- TopHat:** RNA-Seq splice junction mapper
- Cufflinks:** Isoform assembly, quantitation

NIH LIBRARY

<http://bowtie-bio.sourceforge.net/myrna>



Additional help



## Outline

- Databases and tools (free)
- Licensed resources (from the NIH Library)
- Training
- One-on-one consults
- Practical examples

### Licensed Resources for NIH Staff

ArrayStar/QSeq  
CLC Genomics Workbench  
GeneIndexer  
GeneSpring  
Genomatix  
\*\*Human Genome Mutation Database Professional  
Ingenuity Pathways Analysis (IPA)  
Lasergene  
MetaCore™ from GeneGo  
Eureka!  
Open Helix on-line tutorial suite  
Partek Genomics Suite  
ProteinLounge  
SeqMan NGen

### Licensed Resources (from the NIH Library)

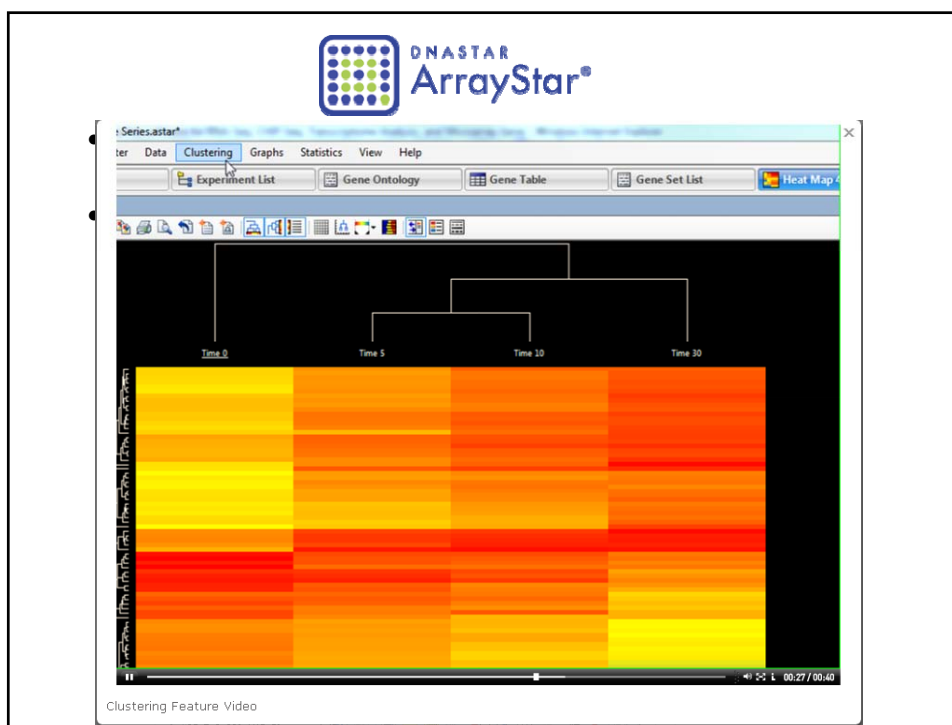
- DNA/protein sequence analysis
- Expression data analysis
- Pathway analysis
- Next gen sequence analysis
- Promoter/SNP prediction





ArrayStar is a gene expression analysis and visualization software on Windows. Tools for Venn diagrams, scatter plot, heat maps and line graphs for clustering.

- Identify relationships between genes with particular biological functions
- Determine relative importance of genes in specific processes using the numerous statistical comparisons
- Download gene ontology tree structure from the Gene Ontology Consortium
- Visualize expression level changes in individual genes over the course of the experiment
- Multi-functional scatter plots generated to easily select groups of genes for analysis
- Normalization methods: RMA, PLIER, quantile normalization, or average summarization
- Cluster data using hierarchical clustering or k-means





## DNA and protein sequence analysis on Windows and Mac Seven applications

- SeqBuilder Sequence editing, annotation, automated virtual cloning, and primer design
- SeqMan Pro Contig assembly and analysis, including SNP discovery, coverage evaluation, and project annotation
- MegAlign DNA and protein sequence alignments and analysis
- GeneQuest Gene discovery and annotation
- Protean Protein structure analysis and prediction
- PrimerSelect Primer design
- EditSeq Importing and editing unusual file types



Selection: 1286086 -> 1286086 = 1

1286060 1286070 1286080 1286090 1286100 1286110 1286120

Translate Consensus GGATAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

DH10B\_v(1396989-3199477) →

GGATAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_93\_262\_396/1 ← GATAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_67\_970\_609/1 ← GATAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_172\_254\_497/2 → GATAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_145\_517\_92/1 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_56\_881\_826/2 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_43\_485\_573/1 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_109\_308\_609/1 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_11\_206\_94 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_72\_255\_86 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_109\_307\_609/1 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_26\_188\_92 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_67\_459\_85 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SLXA-EAS1\_8\_80\_710\_76 → TAAGGCGTTACGCCCGCATCCGGCAATGOTGACCGATGCTGATCGGACG-CTACGGCGTTATCAG

SNP Statistics Report from Contig 'DH10B\_v(1396989 - 3199477)'

All Found SNPs SNPs Summary

Confirmed SNP Putative SNP Rejected SNP Mixed SNP

Show All SNPs Show Counts as a percent

SNP Percent Filter: keep range min. 0 to max. 100.00 Depth 70

10 SNP Columns Rejected: 0 Confirm: 0 Putative: 10 Mixed: 0 Filtered: 62140

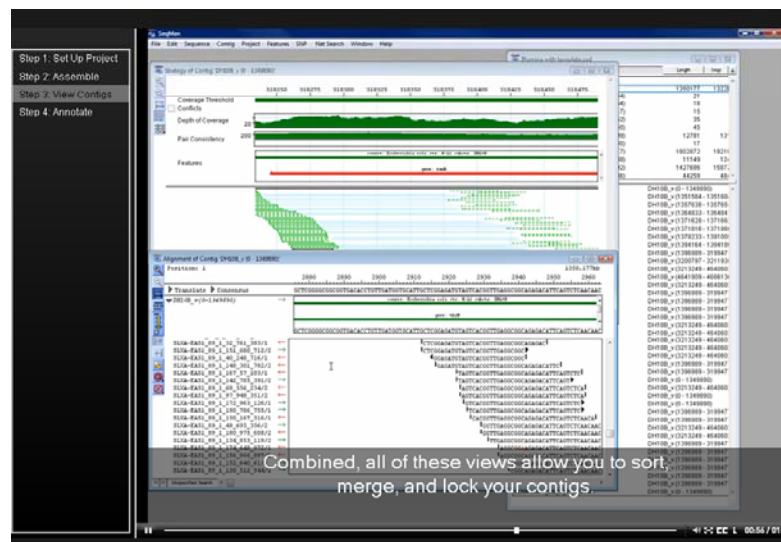
SNP	Contig ID	Contig Pos	Ref Pos	Type
?	DH10B_v(1396989 - 3199477)	1765309	3161929	SNP A
?	DH10B_v(1396989 - 3199477)	1453408	2849800	SNP G
?	DH10B_v(1396989 - 3199477)	1286089	2682809	SNP G
?	DH10B_v(1396989 - 3199477)	1286088	2682808	Indel C
?	DH10B_v(1396989 - 3199477)	1286087	2682807	Indel A
?	DH10B_v(1396989 - 3199477)	1286086	2682806	Indel A
?	DH10B_v(1396989 - 3199477)	1231624	2620356	SNP A
?	DH10B_v(1396989 - 3199477)	1165101	2561651	SNP T
?	DH10B_v(1396989 - 3199477)	1164138	2560888	SNP A
?	DH10B_v(1396989 - 3199477)	1163905	2560655	SNP G

SNP Report Feature Video



SeqMan NGen on Windows and Macintosh is used for traditional, next-generation, and third-generation assemblies.

- Assembly parameter settings are determined for you based on your read technology and your specific project objective
- Integrated with Lasergene's SeqMan Pro for analyzing your project, including discovering SNPs, evaluating coverage, and annotating your consensus sequence
- Next-Gen Workflows Supported
  - *de novo* genome, transcriptome, and metagenome fields of study
  - Whole genome re-sequencing with gap closure, SNP analysis, and annotation
  - Targeted re-sequencing of candidate genes or regions at high accuracy to identify low frequency SNPs and variants
  - Assembly of CHIP-Seq and RNA-Seq reads against a reference genome





QSeq is an application for RNA-Seq, ChIP-Seq, and miRNA alignment and analysis. QSeq is fully integrated with ArrayStar, enabling you to take advantage of its powerful visualization and analytical tools, including using Gene Ontology (GO) annotations for ontology comparisons and gene characterization.

- Select gene sets and export associated reads for sequence assembly, alignment, and detailed analysis
- RNA-Seq analysis to measure gene expression for transcriptomes
- ChIP-Seq peak detection to discover binding sites of DNA-associated proteins
- Easy importation of GO annotations for ontology comparisons and gene characterization.
- High Capacity, High Speed Assembler
  - Align 1 billion or more sequences to a human genome
  - Short and long sequence reads aligned
  - Align 100 million reads in less than one hour on a common desktop machine

The screenshot displays the ArrayStar - QeEST.aster\* application window. The main window shows a 'Peak Table' with a list of genes on the left and a 'Scatter Plot' view in the center. A secondary window titled 'Gene Table' is overlaid, showing a table of gene data. The table has the following columns: DB\_Object\_ID, DB\_Object\_Symbol, DB\_Object\_Synonym, and GO ID. The data rows include various gene symbols and their corresponding GO annotations. A tooltip is visible over the GO ID column for the entry 'SOSS complex, response to DNA damage stimulus', providing additional details about the ontology term.

DB_Object_ID	DB_Object_Symbol	DB_Object_Synonym	GO ID
Q93008	USP9X	USP9X, DFFRX, FAM, USP9, IP100003964, IP...	co-SMAD binding, cysteine-type endopepti...
Q9HC29	NOD2	NOD2, CARD15, IBD1, IP100005559, IP1007...	nucleotide-binding oligomerization domai...
O14949	UQCRCQ	UQCRCQ, IP100024742, QCR8_HUMAN	respiratory chain, mitochondrial inner mem...
Q15628	TRADD	TRADD, IP100018744, TRADD_HUMAN	death domain binding, cytoplasm, kinase bi...
Q81W55	RNF168	RNF168, IP100217899, RNF168_HUMAN	histone H2A K63-linked ubiquitination, ubiq...
Q8TEY7	USP33	USP33, KIAA1097, VDU1, IP100236901, IP10...	protein K63-linked deubiquitination, ubiquit...
Q15018	FAM1758	FAM1758, ABRO1, KIAA0157, IP100299517...	BSIS complex, polyubiquitin binding, prote...
Q724H7	HAUS6	HAUS6, DGT6, FAM29A, KIAA1574, IP10071...	HAUS complex, spindle, mitosis, cell division...
Q9Y4R7	TTL3	TTL3, PRO0207, IP100789856, IP100642191...	protein-glycine ligase activity, initiating, ciliu...
Q9NUN7	ACER3	ACER3, APHC, PHCA, IP100018961, ACER3_...	phytoceramidase activity, membrane endop...
Q8WYB5	MYST4	MYST4, KIAA0383, MORF, MOZ2, IP1000099...	MOZ/MORF histone acetyltransferase com...
Q9NRY2	SSBP1	SSBP1, C9orf80, HSPC043, HSPC291, IP100...	SOSS complex, response to DNA damage stimu...
P08684	CYP3A4	CYP3A4, CYP3A3, IP100465138, CP3A4_HU...	oxidase
P10635	CYP2D6	CYP2D6, CYP2D1, IP100943274, CP2D6_H...	GO ID: response to DNA damage stimuli (GO:0006974)
Q8IU60	DCP2	DCP2, NUDT20, IP100292382, IP100440151...	hist
Q96FW1	OTUB1	OTUB1, OTB1, HSPC263, OTU1, IP100939174...	GO:0006974
Q9Y2K6	USP20	USP20, KIAA1003, LSF3A, VDU2, IP100328...	pro
Q9Y3C0	CCDC53	CCDC53, AD-016, CGI-116, x0009, IP100032...	pro
P52907	CAPZA1	CAPZA1, IP100005969, CAZA1_HUMAN	WA
Q12768	KIAA0196	KIAA0196, IP100029175, STRUM_HUMAN	WA
Q2M389	KIAA1033	KIAA1033, IP100298991, IP100164930, WA...	WA
Q8DM2	C17orf49	C17orf49, IP100790228, CQ049_HUMAN	MLL1 complex, nucleus DNA binding
Q8IZL8	PELP1	PELP1, HMX3, MNAR, IP100006702, IP10078...	MLL1 complex, nucleus transcription, cyto...
Q9V265	BLU/BL1	BLU/BL1, INO80H, NIM238, TR19, TR16A	MLL1 complex, RNA processing, transcrip...

miRNA

mir-29a (Homo sapiens) ATGACTGATTCCTTTGGTGTTCAGAGTCAATAAATTTCTAGCACCATCTGAAATCGGTTAT

Consensus ACTGATTCCTTTGGTGTTCAGAGT\*\*\*\*\*ACTAGCACCATCTGAAATCGGTTAA

Mature\* (2) ACTGATTCCTTTGGTGTTCAG

Mature\* super (1) ACTGATTCCTTTGGTGTTCAG

Precursor variant (1) ACTGATTCCTTTGGTGTTCAGAGT

Mature (30,689) TAGCACCATCTGAAATCGGTTA

Mature sub (1,224) TAGCACCATCTGAAATCGGTT

Mature variant (685) TAGCACCATCTGAAATCGGTA

Mature super variant (377) TAGCACCATCTGAAATCGGTTAA

Mature variant (241) TAGCACCATCTGAAATCGGTT

Mature super (240) CTAGCACCATCTGAAATCGGTTA

**Alignment against the mir29a precursor sequence**

**A Table of Deletions and Insertions**

- *de novo* Assembly on a strong desktop computer
- Analysis of NGS data
- RNA-Seq, ChIP-Seq
- mircoRNA analysis
- SNP and DIP detection

Refere...	Refe...	Variants	Allele var...	Frequencies	Counts	Coverage	Overlappin...	Amino acid ...
47	A	2	A/-	70,6/29,4	12/5	17		
11956	T	2	T/-	75,7/24,3	14/5	16	19	Gene: yash1, ... Change, fra...
29376	T	2	T/-	56,2/43,8	9/7	16		
34112	T	2	-/T	50,0/50,0	11/11	22		
43091	A	2	A/-	77,3/22,7	17/5	22	22	Gene: ftd, C, ... Change, fra...
81649	T	1	-	71,4	10	14		
93816	-	2	-/T	43,6/56,4	7/4	11	11	Gene: ampE, ... Change, fra...
93943	G	2	G/-	69,0/31,2	11/5	16	16	Gene: ampE, ... Change, fra...
101718	-	2	-/T	78,3/21,7	16/5	23		
103493	T	2	T/-	70,0/30,0	14/5	20		
108903	GCAT	2	GCAT/----	79,2/20,8	19/5	24	24	Gene: speD, ... Change, fra...
119075	A	2	A/-	68,4/31,6	13/6	19	19	Gene: ygd, ... Change, fra...
121237	CTGA	2	CTGA/----	72,0/28,0	18/7	25	25	Gene: ygd, ... Change, fra...
123650	-	2	-/T	76,0/24,0	19/6	25	25	Gene: parE, ... Change, fra...
123655	T	2	T/-	84,2/15,8	13/11	24	24	Gene: parE, ... Change, fra...
130976	T	2	T/-	70,6/29,4	12/5	17	17	Gene: ygd, ... Change, fra...

## GeneSpring GX

### Platform Strengths

Powerful Statistical Tools and Sophisticated Data Visualizations

Biological Contextualization (pathway module, GSEA, GSA, GO analysis)

Open Platform  
Agilent, Affymetrix, Illumina, custom PCs, Macs, Linux

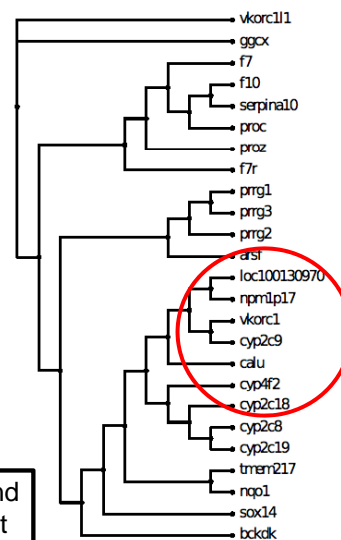
# GeneIndexer

Discovery and hypotheses generation tool that compresses literature analysis from months to minutes

- Extracts implicit and explicit gene associations from over 1.5 million scientific literature abstracts
- Use free text queries (diseases, pathways, phenotypes, drugs, GO classifications, etc.) to identify and prioritize genes most relevant to a given research question.
- Builds hierarchical trees in which genes are clustered into functionally related groups.

## Case Study: A Warfarin Query

Rank	Gene	Score
1	vkorc1f1	0.752
2	loc100130970	0.647
3	npmlp17	0.647
4	vkorc1	0.597
5	cyp2c9	0.572
6	calu	0.553
7	ggcx	0.513
8	cyp2c19	0.454
9	f10	0.424
10	prng1	0.414
11	cyp2c18	0.400



GeneIndexer quickly finds CYP2C9 and VKORC1, plus other genes of interest

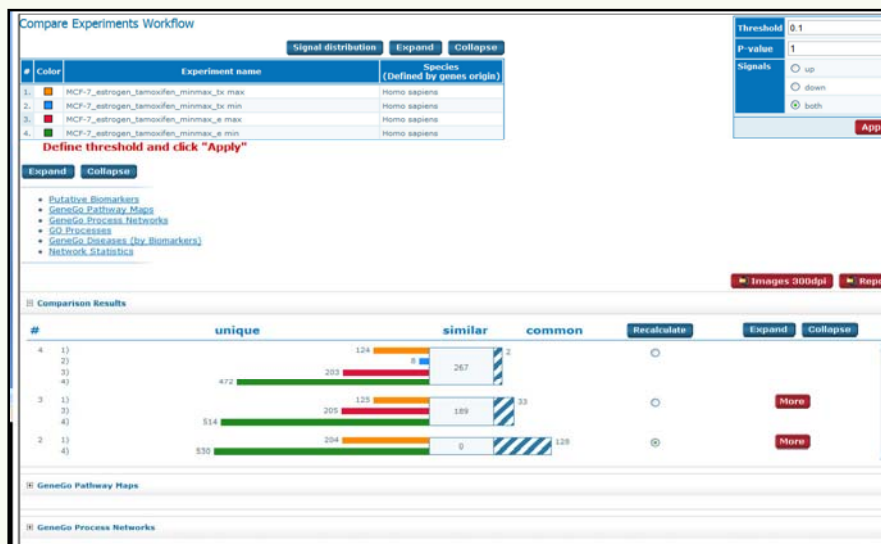


## MetaCore from GeneGo

- Analyze data such as microarray gene expression, SNPs, metabolic profiles, high content screening (HCS) assays
- Identify the most relevant pathways, networks and cellular processes.
- Search information about genes, proteins, compounds, pathway maps and diseases

<http://www.genego.com/metacore.php>

## MetaCore from GeneGo



<http://www.genego.com/metacore.php>

## Eureka!

- Search GeneGo's knowledge database
- Find pathways and enrichments for biomarkers, signaling and metabolic processes

<http://www.genego.com/eureka.php>

## Eureka!

The screenshot displays the Eureka! search results for the query 'akt1'. The interface includes a search bar with 'akt1' entered, a search button, and an 'Exact match' checkbox. Below the search bar, there are navigation links for 'Advanced Search' and 'Search compounds by structure'. The results are organized into three main sections:

- Objects Found:** A sidebar on the left lists various object types: Genes (19), Proteins (17), Compounds (228), Drugs (1), Small Molecule Drugs (1), Network Objects (17), Interactions (544), GeneGo Maps (183), GeneGo Networks (235), and Articles (401). Below this is a 'Selected GeneGo Maps' section.
- Result pages:** A navigation bar showing pages 1 through 10, with 'Next' and '(Showing results 1 to 10 of 181)'.
- Search Results:** Three results are listed:
  - Development\_EDG1 signaling pathway:** Includes a network diagram and a description: "Network Objects: G-protein alpha-12, L-Arginine, cytochrome, G-protein alpha-13, Ca<sup>2+</sup>, endothelin, differentiation gene 1 (EDG1). Description: EDG1 signaling pathway Sphingosine-1-phosphate receptor 1, also known as Endothelial differentiation gene 1 (EDG1) is a high affinity receptor for the bioactive lipid Sphingosine-1-phosphate. Although the EDG1 expression is very abundant in endothelial cells, transcripts related to EDG1 are also detected at lower levels in vascular smooth muscle cells, fibroblasts, melanocytes, cells of epithelial origin, brain alveolar macrophages cardiovascular tissues natural killer cells and T cells. EDG1 is a G-protein coupled receptor which interacts with Guanine nucleotide binding protein alpha inhibiting activity polypeptide 1, 2 and 3 [ G-protein alpha-1, G-protein alpha-2 and G-protein alpha-2 ], and Guanine nucleotide binding protein, alpha activating activity polypeptide G ( G-protein alpha-0 ), G-protein alpha-i family members inhibit Adenylate cyclase activity. Thereby, binding of..."
  - IGF-1 signaling in pancreatic cancer:** Includes a network diagram and a description: "Network Objects: SMYD3, alpha 1 subunit, IGF-1, MMP-14, AT5, GRK2, IGF-1 receptor, JAK1, MEK1. Description: IGF-1 signaling in pancreatic cancer Abstract IGF-1 signaling is enhanced in pancreatic cancer cells where it regulates cancer growth and progression. Pancreatic cancer cells overexpress IGF-1 receptor, IRS-1 and IRS-2. IGF-1 binding to IGF-1 receptor activates FAK1, PI3K/AKT, ERK1/2, JAK1, JAK2/ STAT3 pathways leading to overexpression of VEGF-A, COX-2 (PTGS2) and activation of Src, Src-2 and Arp2/3 complex. Thereby, IGF-1 signaling regulates survival of pancreatic cancer cells and increases angiogenesis and invasion. Details Insulin-like growth factor 1 ( IGF-1 ) signaling is required for maintenance of growth and tumorigenicity of pancreatic cancer regulating invasion, angiogenesis and survival of pancreatic cancer cells. In comparison to normal pancreas, pancreatic adenocarcinomas overexpress the Insulin-like growth factor 1 receptor ( IGF-1 receptor )..."
  - KLF6 and regulation of KLF6 alternative splicing in HCC:** Includes a network diagram and a description: "Network Objects: SFRS1 (SFP1), KLF6, Srs, CDK6, EGR3, rhltna(4.5)PC intracellular, H-Ras, SDC..."

At the bottom of the screenshot, the text 'MetaCore™ version 6.1 build 23116' and 'Copyright © 2000-2009 GenGo Inc.' are visible.

<http://www.genego.com/eureka.php>

**genomatix software suite** v2.0

Main menu | Logout |

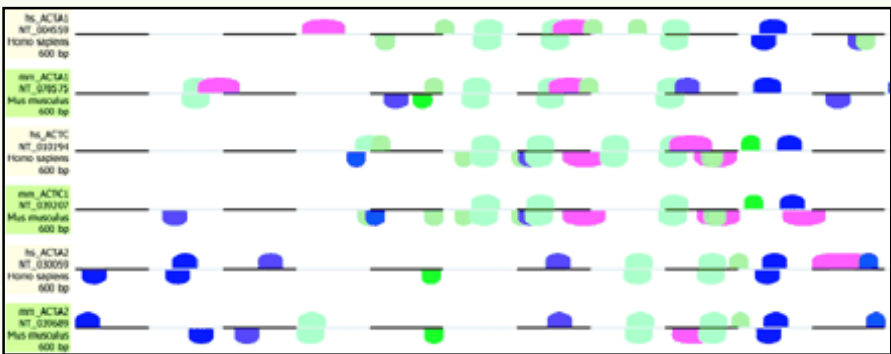
NGS Analysis Literature & Pathways Genomes & Data Pattern Search & Analysis Pattern Definition Alignment & Mapping Tools Projects & Account Help

- Analysis of genomic data, gene regulation and expression
- Generate and evaluate networks and pathways
- Perform literature searches and sequence analyses
- Visualize comprehensive genome annotation

**genomatix software suite** v2.0

Main menu | Logout |

NGS Analysis Literature & Pathways Genomes & Data Pattern Search & Analysis Pattern Definition Alignment & Mapping Tools Projects & Account Help




TFBS common to all sequences in a set of six muscle actin promoters

Cartharius et al. *Bioinformatics* 21(13):2933-42

Coming Soon.....

HGMD Professional



**BIOBASE**  
BIOLOGICAL DATABASES

• Home • Contact • Sitemap

You are here : Products / HGMD® Professional

▼ Products

- > BKL PROTEOME
- > BKL TRANSFAC
- > Explain Analysis System
- ▼ HGMD® Professional
  - HGMD Benefits
  - Example Results
  - Statistics
  - References
  - FAQ
- > BRENDA Professional
- > CLC Genomics Workbench
- > Services
- > Buy

HGMD® - Human Gene Mutation Database

Introducing disease mutations to genetics and genomics research

Coming Soon.....

HGMD Professional

Applications of HGMD® include:

- \* Determining the novelty of identified gene mutations
- \* Obtaining all known mutations for a given gene or all mutated genes for a disease of interest
- \* Mapping mutations to a full genome sequence
- \* Location in a particular motif within a splice site or regulatory region



## Ingenuity Pathway Analysis

- Search and Explore Biological and Chemical Knowledge  
Genes, drugs, chemicals, protein families, normal cellular and disease processes, and signaling and metabolic pathways
- Dynamic Signaling & Metabolic Pathways
- Analyze and Interpret Data  
such as gene expression, SNP microarrays, proteomics experiments, and gene lists

[http://www.ingenuity.com/products/pathways\\_analysis.html](http://www.ingenuity.com/products/pathways_analysis.html)

## Ingenuity Pathway Analysis

Summary | Networks | Functions | Canonical Pathways | My Pathways | Gene Summary | Network Explorer | Overlapping Networks | Lists

Here is a summary of analysis Waring\_tox\_dataset - 2007-01-30 08:20 PM


Top Networks			
ID	Associated Network Functions	Score	Preview
1 View	Cell Death, Cell Cycle, Cancer	94	Preview
2 View	Cell Death, Gene Expression, Cellular Development	30	Preview
3 View	Cell Death, Cancer, Cellular Growth and Proliferation	27	Preview
4 View	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry	23	Preview
5 View	DNA Replication, Recombination, and Repair, Cellular Growth and Proliferation, Cancer	11	Preview

Top Tox Functions			
Hepatotoxicity	Name	Significance	# Genes
	Liver Necrosis/Cell Death	3.81E-7 - 4.60E-6	5
	Liver Hyperplasia/Hyperproliferation	8.38E-6 - 8.38E-6	3

Top Toxicity Lists		
Name	Ratio	Significance
Aryl Hydrocarbon Receptor Signaling	3/9 (0.33333)	5.79E-3
Cytochrome P450 Panel - Substrate is an Eicosanoid (Rat)	1/3 (0.33333)	1.27E-1
Cytochrome P450 Panel - Substrate is a Xenobiotic (Human)	5/19 (0.26316)	1.07E-3
Cytochrome P450 Panel - Substrate is a Xenobiotic (Rat)	6/23 (0.26087)	3.42E-4
Cytochrome P450 Panel - Substrate is a Xenobiotic (Mouse)	6/28 (0.21429)	1.07E-3

Top Canonical Pathways		
Top Genes		
Normalized Ratio up-regulated		
Gene	Exp. Value	Exp. Chart
CYP1A1	↑123.700	--
ABCB1B	↑100.100	--
COX5C (includes EG54322)	↑29.100	--
ABCC3	↑26.900	--
GGPD	↑26.600	--
JUNB	↑21.600	--
GGT1	↑17.800	--

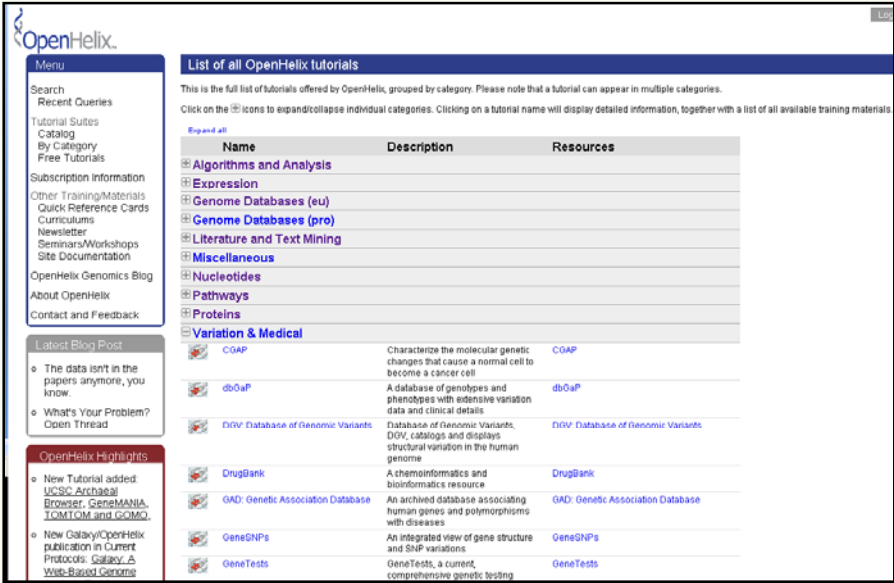
[http://www.ingenuity.com/products/pathways\\_analysis.html](http://www.ingenuity.com/products/pathways_analysis.html)



# OpenHelix

- Search function to find the right bioinformatics resource
- Tutorial Suites to learn the resource
  - Over 100 on-line self-run narrated tutorials
  - Includes training materials—Powerpoint Slides, Handouts and Exercises
  - Covering a range of needs and research areas
  - Continually updated as resources change
- IP based access at [www.openhelix.com](http://www.openhelix.com)

Copyright OpenHelix. No use or reproduction without express written consent 43



**OpenHelix**

**Menu**

- Search
- Recent Queries
- Tutorial Suites
- Catalog
- By Category
- Free Tutorials
- Subscription Information
- Other Training/Materials
- Quick Reference Cards
- Curriculums
- Newsletter
- Seminars/Workshops
- Site Documentation
- OpenHelix Genomics Blog
- About OpenHelix
- Contact and Feedback


**Latest Blog Post**

- o The data isn't in the papers anymore, you know.
- o What's Your Problem? Open Thread


















**OpenHelix Highlights**

- o New Tutorial added: UCSC Archival Browser, GeneMANIA, TOMTOM and GOMO.
- o New Galaxy/OpenHelix publication in Current Protocols: Galaxy: A Web-Based Course


**List of all OpenHelix tutorials**

This is the full list of Tutorials offered by OpenHelix, grouped by category. Please note that a tutorial can appear in multiple categories. Click on the  icons to expand/collapse individual categories. Clicking on a tutorial name will display detailed information, together with a list of all available training materials.

[Expand all](#)


Name	Description	Resources
 <a href="#">Algorithms and Analysis</a>		
 <a href="#">Expression</a>		
 <a href="#">Genome Databases (eu)</a>		
 <a href="#">Genome Databases (pro)</a>		
 <a href="#">Literature and Text Mining</a>		
 <a href="#">Miscellaneous</a>		
 <a href="#">Nucleotides</a>		
 <a href="#">Pathways</a>		
 <a href="#">Proteins</a>		
 <a href="#">Variation &amp; Medical</a>		
 <a href="#">COAP</a>	Characterize the molecular genetic changes that cause a normal cell to become a cancer cell	<a href="#">COAP</a>
 <a href="#">dbGaP</a>	A database of genotypes and phenotypes with associated variation data and clinical details	<a href="#">dbGaP</a>
 <a href="#">DIV: Database of Genomic Variants</a>	Database of Genomic Variants, DIV, catalogs and displays structural variation in the human genome	<a href="#">DIV: Database of Genomic Variants</a>
 <a href="#">DrugBank</a>	A cheminformatics and bioinformatics resource	<a href="#">DrugBank</a>
 <a href="#">GAD: Genetic Association Database</a>	An archived database associating human genes and polymorphisms with diseases	<a href="#">GAD: Genetic Association Database</a>
 <a href="#">GeneSNPs</a>	An integrated view of gene structure and SNP variations	<a href="#">GeneSNPs</a>
 <a href="#">GeneTests</a>	GeneTests, a current, comprehensive genetic testing	<a href="#">GeneTests</a>

## Partek GS™ for Integrated Genomics



**Microarray**

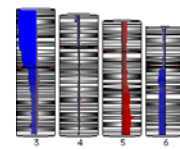
&



**Next Generation Sequencing**

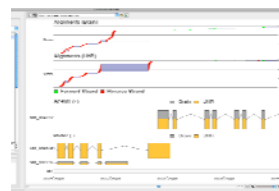
**Genome**

- Copy Number
  - Total & Allele Specific
- Association
- Loss of Heterozygosity



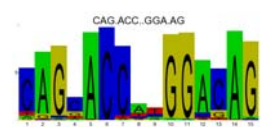
**Transcriptome**


- Gene Expression
- Exon/Alternative Splicing
- DGE & mRNA -Seq



**Regulation**

- ChIP-Chip
- ChIP-Seq
- microRNA





45

Copyright © Partek Inc.

# ProteinLounge<sup>BETA</sup>

Site Best Viewed in  
Register |

Webtop
My BioLife
BioShare
Reagents
Pathways
Pathway Builder
Tools

Brought to you by National Institutes of Health

Login

Alert

Friends


People

Favorites

Open Items

Feedback

Help



Advertise with us

www.proteinlounge.com

BioShare

watch and share **Biological articles**  
videos/animations images and **many more...**

Place your AD here

ENJOY the


What's NEW!

Online PathwayBuilder

- DRAW
- SAVE
- SHARE
- PUBLISH

Is now ONLINE

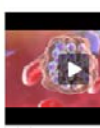
Featured Pathways



BAFF in B-Cell Signaling

Members of the **TNF** (Tumor Necrosis Factor) ligand family play important roles in various physiological and pathological processes, including cell proliferation, differentiation, apoptosis, modulation of immune response, and induction of inflammation. The **BAFF** (B-Cell Activating Factor) [also known as **BLYS** (B-Lymphocyte Stimulator), **TALL** (TNF- and Activates Apoptosis, NF- $\kappa$ B $\beta$ , and JNK) or **THANK** (TNF Homolog that Activates B-Cell survival, **BAFF** is expressed on dendritic cells, monocytes/macrophages, and T-Cells. [more](#)

BioAnimations



Malaria

Malaria is an infectious disease caused by a protozoan parasite of the genus *Plasmodium*, which is transmitted to human by the bite of infected female *Anopheles* mosquitoes. The life cycle of *Plasmodium* inside human body begins when the *Plasmodium* sporozoites from female *Anopheles*.

110 views

ProteinLounge Animation

Biotech Jobs

Settings

Postdoctoral Positions 23 minutes ago

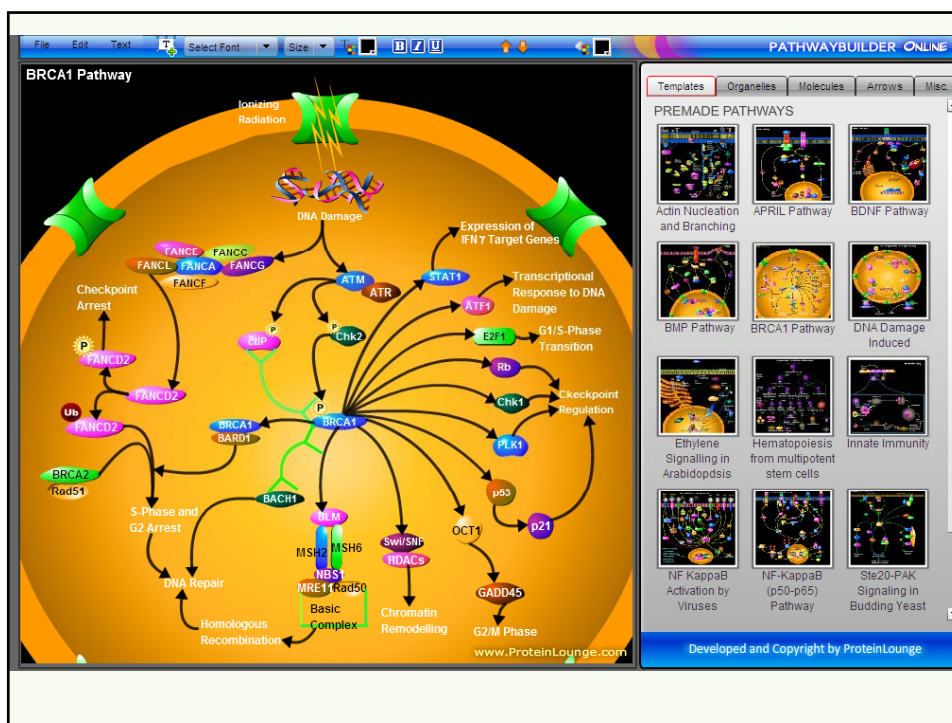
Tenure Eligible Faculty Position 47 minutes ago

Post Doctoral Fellowships in Transdisciplinary Research into Plasma White

Biotech News

Settings

Transcript Partners Continues Expansion with Appointment of Gareth Walters



## Licensed Resources (from the NIH Library)

- DNA/protein sequence analysis
- Expression data analysis
- Pathway analysis
- Next gen sequence analysis
- Promoter/SNP prediction



## Licensed Resources (from the NIH Library)

Floating network seat(s)

to register go to  
<http://nihlibrary.nih.gov/bioinformatics>

Some licensed resources loaded on two  
dedicated computers in the Library

additional software such as

Cytoscape  
Cn3D

## Outline

- Databases and tools (free)
- Licensed resources (from the NIH Library)  
Dedicated computers
- Specific examples
- Training and additional help

## Practical examples

- Obtain human non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins
- Identify SNPs in conserved regions
- Download upstream sequences for multiple human genes
- Obtain unique genes in a genome compared to other genome(s)



## Practical examples

- Non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins (NCBI)
- SNPs in the conserved regions (UCSC genome browser)
- Download upstream sequences for multiple human genes (Ensembl)
- Obtain unique genes in a genome compared to other genome(s) (IMG)



## Non-synonymous human SNPs that are associated with diseases and with known 3-D structures of the proteins

The screenshot shows the NCBI homepage with a search bar at the top. A purple arrow points to the 'Search' button. The page includes a 'resources' sidebar on the left, a 'Welcome to NCBI' message, a 'Genome' section with a '1000 prokaryotic genomes are now completed and available in the Genome database.' announcement, and a 'Popular Resources' section on the right.

<http://www.ncbi.nlm.nih.gov/>

## Non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins

The screenshot shows the Entrez cross-database search page. A purple arrow points to the 'SNP: single nucleotide polymorphism' entry in the list of databases. The page includes a search bar at the top and a grid of database icons and descriptions.

NCBI ENTREZ SNP Single Nucleotide Polymorphism

All Databases Published Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search SNP for [ ] Go Clear

Limits Preview/Index History Clipboard Details

Click on the image below to view the connections between Entrez SNP and other databases.

dbSNP BUILD 131

Entrez SNP  
Search SNP  
Search Mouse SNP  
Common Query Filters  
Entrez Batch Query  
SNP Link Data model

My NCBI  
My NCBI help

Entrez SNP Help  
Searchable FAQ  
Search Fields  
Programming Utilities  
Batch Report  
Legend  
Examples  
dbSNP Home Page  
Overview

Entrez Help  
General help  
Limits

SNP

LIBRARY  
NIH

dbSNP BUILD 131

Limit your search by any of the following criteria.

<p><b>Organism</b> CLEAR</p> <input type="checkbox"/> Bos taurus <input type="checkbox"/> Caenorhabditis elegans <input type="checkbox"/> Canis familiaris <input type="checkbox"/> Danio rerio <input type="checkbox"/> Felis catus <input type="checkbox"/> Gallus gallus <input type="checkbox"/> Equus caballus <input checked="" type="checkbox"/> Homo sapiens <input type="checkbox"/> Macaca mulatta <input type="checkbox"/> Monodelphis domestica	<p><b>Chromosomes</b> CLEAR</p> <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 2a <input type="checkbox"/> 2b <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> X <input type="checkbox"/> Y	<p><b>Chromosome Range</b> CLEAR</p> <p>From: [ ] To: [ ]</p>
<p><b>Map Weight</b> CLEAR</p> <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3-10 <input type="checkbox"/> 10+	<p><b>Function Class</b> CLEAR</p> <input checked="" type="checkbox"/> coding nonsynonymous <input checked="" type="checkbox"/> nonsense <input checked="" type="checkbox"/> missense <input checked="" type="checkbox"/> frame shift <input type="checkbox"/> intron <input type="checkbox"/> coding synonymous <input type="checkbox"/> locus region <input type="checkbox"/> 5' utr <input type="checkbox"/> 3' utr	<p><b>SNP Class</b> CLEAR</p> <input type="checkbox"/> het <input type="checkbox"/> in del <input type="checkbox"/> microsatellite <input type="checkbox"/> mixed <input type="checkbox"/> mnp <input type="checkbox"/> named <input type="checkbox"/> no variation <input type="checkbox"/> snp
<p><b>Method Class</b> CLEAR</p> <input type="checkbox"/> computed <input type="checkbox"/> by-cluster	<p><b>Validation Status</b> CLEAR</p> <input type="checkbox"/> by-cluster	<p><b>Variation Allele</b> CLEAR</p> <input type="checkbox"/> A

<p><b>Annotation</b> CLEAR</p> <input type="checkbox"/> Clinical/USDB Submissions <input type="checkbox"/> nucleotide <input checked="" type="checkbox"/> OMIM <input type="checkbox"/> protein <input type="checkbox"/> structure <input type="checkbox"/> PubMed <input checked="" type="checkbox"/> Cited in PubMed	<p><b>Heterozygosity</b> CLEAR</p> <input type="checkbox"/> 0-10 <input type="checkbox"/> 40-50 <input type="checkbox"/> 10-20 <input type="checkbox"/> 20-30 <input type="checkbox"/> 30-40	<p><b>Success Rate</b> CLEAR</p> <input type="checkbox"/> 80-85 <input type="checkbox"/> 85-90 <input type="checkbox"/> 90-95 <input type="checkbox"/> 95+
--	--	---

<p><b>Created Build ID</b> CLEAR</p> <input type="checkbox"/> Current build ID(131) <input type="checkbox"/> Last Build ID(130)	<p><b>Updated Build ID</b> CLEAR</p> <input type="checkbox"/> Current build ID(131) <input type="checkbox"/> Last Build ID(130)	<p><b>Individual SNP</b> CLEAR</p> <input type="checkbox"/> Venter <input type="checkbox"/> Watson <input type="checkbox"/> Chinese_YH1
--	--	---

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search SNP for  Go Clear Save Search

Limits Preview/Index History Clipboard Details

Limits: **snp\_omim, snp\_structure, snp\_pubmed\_cited, nonsense, missense, frameshift, homo sapiens**

Display Graphic Summary Show 20 Sort By Send to

All: 19 1000 Genomes: 13 Cited in PubMed: 19 Clinical/LSDB Submissions: 16 Human: 19

Items 1 - 19 of 19 One page.

1: rs1343879 [*Homo sapiens*] Links

GATCTCAGGGAAGTCTGAGTACT[**A/C/G/T**]TCTGAGAACTCCAGCATCTCGGAC ABI, ILLUMINA, TSC-CSHL

MapView No VarVu PubMed GeneView SeqView Protein 3D OMIM

HGVSNames: [NM\_138703.3:c.358G>A] [NM\_138703.3:c.358G>C] [NM\_138703.3:c.358G>T] [NP\_619648.1:p.Glu120Gln] [NP\_619648.1:p.Glu120Lys] [NP\_619648.1:p.Glu120X] [NT\_011669.17:g.13322517C>A] [NT\_011669.17:g.13322517C>T]

2: rs5030865 [*Homo sapiens*] Links

TTTGTGCCCTTCTGCCATCACCCAC[**A/C/G/T**]GGAAGTGGTTGGCGAAGGGGCACAA HGBASE

MapView VarVu PubMed GeneView SeqView Protein 3D OMIM

HGVSNames: [NG\_000853.3:g.1924G>A] [NG\_000853.3:g.1924G>C] [NG\_000853.3:g.1924G>T] [NG\_003180.2:g.26484G>A] [NG\_003180.2:g.26484G>C] [NG\_003180.2:g.26484G>T] [NG\_008376.1:g.6849G>A] [NG\_008376.1:g.6849G>C] [NG\_008376.1:g.6849G>T] [NM\_000106.4:c.505G>A] [NM\_000106.4:c.505G>C] [NM\_000106.4:c.505G>T] [NM\_001025161.1:c.353-89G>A] [NM\_001025161.1:c.353-89G>C] [NM\_001025161.1:c.353-89G>T] [NP\_000097.2:p.Gly169Arg] [NP\_000097.2:p.Gly169X] [NT\_011520.12:g.21915604C>A] [NT\_011520.12:g.21915604C>G] [NT\_011520.12:g.21915604C>T]

NCBI Resources How To My NCBI Sign In

PubMed.gov Search: PubMed Limits Advanced search Help

U.S. National Library of Medicine National Institutes of Health

Search Clear

Display Settings: Abstract Send to

Full text free on BioMed Central FREE full text article in PubMed Central

BMC Med Genomics, 2008 Jun 11;1:24.

**Genotyping panel for assessing response to cancer chemotherapy.**

Dai Z, Papp AC, Wang D, Hampel H, Sadee W.

Program in Pharmacogenomics, Department of Pharmacology, Comprehensive Cancer Center, College of Medicine and Public Health, The Ohio State University, 5072 Graves Hall, 333 West 10th Avenue, Columbus, OH 43210-1239, USA. dai.15@osu.edu.

**Abstract**

**BACKGROUND:** Variants in numerous genes are thought to affect the success or failure of cancer chemotherapy. Interindividual variability can result from genes involved in drug metabolism and transport, drug targets (receptors, enzymes, etc), and proteins relevant to cell survival (e.g., cell cycle, DNA repair, and apoptosis). The purpose of the current study is to establish a flexible, cost-effective, high-throughput genotyping platform for candidate genes involved in chemoresistance and -sensitivity, and treatment outcomes.

**METHODS:** We have adopted SNPlex for genotyping 432 single nucleotide polymorphisms (SNPs) in 160 candidate genes implicated in response to anticancer chemotherapy.

**RESULTS:** The genotyping panels were applied to 39 patients with chronic lymphocytic leukemia undergoing flavopiridol chemotherapy, and 90 patients with colorectal cancer. 408 SNPs (94%) produced successful genotyping results. Additional genotyping methods were established for polymorphisms undetectable by SNPlex, including multiplexed SNaPshot for CYP2D6 SNPs, and PCR amplification with fluorescently labeled primers for the UGT1A1 promoter (TA)nTAA repeat polymorphism.

**CONCLUSION:** This genotyping panel is useful for supporting clinical anticancer drug trials to identify polymorphisms that contribute to interindividual variability in drug response. Availability of population genetic data across multiple studies has the potential to yield genetic biomarkers for optimizing anticancer therapy.

**Related citations**

Genotyping panel for assessing response to cancer chemotherapy [BMC Med Genomics. 2008]

Validation of the performance of a comprehensive genotyping array [Hum Mutat. 2009]

GMFilter and SXTestPlate: software tools for improving the SNPlex [BMC Bioinformatics. 2009]

Review Can UGT1A1 genotyping reduce morbidity and mortality in pain? [Genet Med. 2009]

Review Single nucleotide polymorphisms in DNA repair genes and [Methods Mol Biol. 2009]

See reviews...  
See all...

**All links from this record**

Related Citations

Gene

Gene (GeneRIF)

HomoloGene

Nucleotide (RefSeq)

**.0004 DEBRISOQUINE, POOR METABOLISM OF [CYP2D6, GLY169TER]** [dbSNP:rs5030865](#)

In a Caucasian patient with deficiency of the CYP2D6 enzyme and poor metabolism (608992), Broly et al. (1995) identified a gly169-to-ter (G169X) mutation in exon 3 of the CYP2D6 gene.

**.0005 DEBRISOQUINE, POOR METABOLISM OF [CYP2D6, PRO34SER]** [dbSNP:rs106582,72552269](#)

This allelic variant is also known as CYP2D6\*10 or CYP2D6(J) or CYP2D6(CH1, CH2).

Kagimoto et al. (1999) identified a 188C>T transition in exon 1 of the CYP2D6 gene, resulting in a pro34-to-ser (P34S) substitution as a cause of the debrisoquine poor metabolizer phenotype (608992). Takamura et al. (2002) presented data strongly suggesting that catalytic activity as well thermal stability of the enzyme is affected by the P34S polymorphism. They proposed that thermal instability together with reduced intrinsic clearance of CYP2D6\*10 is one reason for the finding of lower metabolic activities for drugs metabolized by CYP2D6 in Orientals, who have a high frequency of CYP2D6\*10, compared with Caucasians.

**.0006 DEBRISOQUINE, POOR METABOLISM OF [CYP2D6, 1-BP DEL, 2637A]**

This allelic variant is also known as CYP2D6\*3 or CYP2D6(A).

Marez et al. (1997) identified a 1-bp deletion (2637A) in exon 3 of the CYP2D6 gene in a group of individuals with the poor metabolizer phenotype (608992).

**.0007 DEBRISOQUINE, ULTRARAPID METABOLISM OF [CYP2D6, ARG296C]**

This allelic variant is also known as CYP2D6\*2 or CYP2D6L.

In a family in which 2 sibs and their father had MRs of less than 0.02 (ultrarapid) (1292) found 13 extra copies of the CYP2D6 gene inherited in an autosomal dominant manner. The authors found 2 extra copies of the CYP2D6 gene, termed CYP2D6L, which contained 2 amino acid substitutions: a 2 arg296-to-cys (R296C), and a 4268G-to-C transversion in exon 9, resulting in a substitution. The MR of individuals with 1 copy of the CYP2D6L gene did not differ from that of individuals with 1 copy of the wildtype. Residue 296 falls within a presumed substrate recognition site, and is a binding site.

**MIM ID +124030**  
CYTOCHROME P450, SUBFAMILY IID, POLYPEPTIDE 6; CYP2D6

**Allelic Variants (Selected Examples) Notes**

Number	Phenotype	Mutation	dbSNP
.0001	DEBRISOQUINE, POOR METABOLISM OF	CYP2D6, IVSD55, G-A, +1	-
.0002	DEBRISOQUINE, POOR METABOLISM OF	CYP2D6, DEL	-
.0003	DEBRISOQUINE, POOR METABOLISM OF	CYP2D6, 1-BP DEL, 1795T	-
.0004	DEBRISOQUINE, POOR METABOLISM OF	CYP2D6, GLY169TER	<a href="#">rs5030865</a>
.0005	DEBRISOQUINE, POOR METABOLISM OF	CYP2D6, PRO34SER	<a href="#">rs106582,72552269</a>
.0006	DEBRISOQUINE, POOR METABOLISM OF	CYP2D6, 1-BP DEL, 2637A	-
.0007	DEBRISOQUINE, ULTRARAPID METABOLISM OF	CYP2D6, ARG296CYS AND SER486THR	<a href="#">rs16247,1133840</a>
.0008	CODEINE, ULTRARAPID METABOLISM OF	CYP2D6, DUP	-

Skip to main content

**Single Nucleotide Polymorphism**

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search for SNP on NCBI Reference Assembly

Search Entrez SNP for Go

**BUILD 131**  
Have a question about dbSNP? Try searching the SNP FAQ Archive!

To display 3D structure, download Ch3D viewer

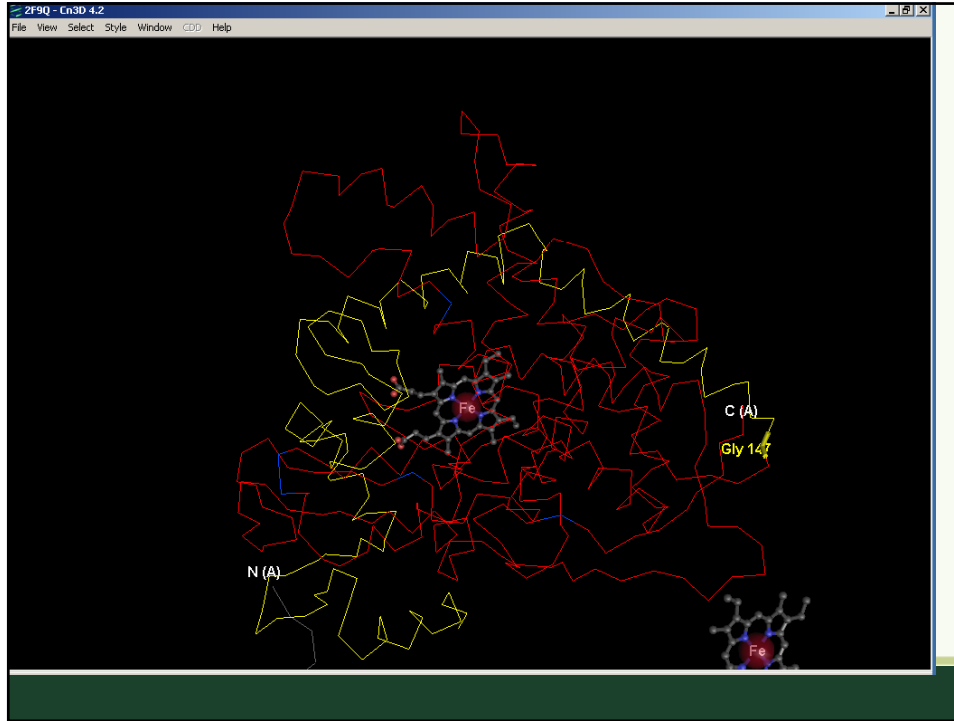
**136 variations mapped to protein NP**

PDB Blast Result: [View Alignment](#)

Protein: [gi40805836](#) 1-1 cytochrome P450 2D6 isoform 1 [Homo sapiens]

Structure Neighbor: [gi85544647](#) 1-1 Chain A, Crystal Structure Of Human Cytochrome P450 2D6

SNP ID	Position protein	Amino Acid	dbSNP	Amino Acid	Position neighbor	CADD
rs79802111	50	ASP	ASP	ASP	26	0.0
rs28371703	91	LEU	LEU	LEU	69	0.0
rs28371703	91	LEU	MET	LEU	69	0.0
rs28371703	91	LEU	VAL	LEU	69	0.0
rs28371704	94	VAL	ASP	VAL	72	0.0
rs76802407	97	ASP	GLU	ASP	75	0.0
rs28371705	98	THR	THR	THR	76	0.0
rs76197528	104	VAL	ALA	VAL	82	0.0
rs28371706	107	THR	ILE	THR	85	0.0
rs28371706	107	THR	ASN	THR	85	0.0
rs28371706	107	THR	SER	THR	85	0.0
rs74802369	107	THR	SER	THR	85	0.0
rs78459009	109	ILE	VAL	ILE	87	0.0
rs1135821	111	GLY	GLY	GLY	89	0.0
rs111519167	111	GLY	SER	GLY	89	0.0
rs1091003	112	PHE	PHE	PHE	90	0.0
rs76060075	118	GLY	ALA	GLY	96	0.0
rs1135822	120	PHE	ILE	PHE	98	0.0
rs61736507	120	PHE	LEU	PHE	98	0.0
rs1135823	122	ALA	SER	ALA	100	0.0
rs60019788	125	PRO	THR	PRO	104	0.0
rs1058164	136	VAL	VAL	VAL	114	0.0
rs61736512	136	VAL	PHE	VAL	114	0.0
rs61736512	136	VAL	ILE	VAL	114	0.0
rs61736512	136	VAL	LEU	VAL	114	0.0
rs61736911	136	VAL	VAL	VAL	114	0.0
rs78482788	151	GLN	GLU	GLN	129	0.0
rs28371710	155	GLU	LYS	GLU	133	0.0
rs28371710	155	GLU	GLN	GLU	133	0.0
rs28371710	155	GLU	GLU	GLU	133	0.0
rs74962936	157	ALA	ALA	ALA	135	0.0
rs1135824	166	ASN	ASP	ASN	144	0.0
rs1135825	167	HIS	HIS	HIS	145	0.0
rs1135826	167	HIS	GLN	HIS	145	0.0
rs1135826	168	SER	ALA	SER	146	0.0
rs1135826	168	SER	PRO	SER	146	0.0
rs1135826	168	SER	THR	SER	146	0.0
rs6030866	169	GLY	GLY	GLY	147	0.0
rs67780109	173	ARG	CYS	ARG	161	0.0
rs72649356	174	PRO	PRO	PRO	152	0.0
rs2267449	176	GLY	GLY	GLY	154	0.0



## To download SNPs in conserved regions

**Table Browser**

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For more information, see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological annotations, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data.

clade:  genome:  assembly:

group:  track:

table:

on:  genome  position

identifiers (names/accession):

f 2

intersection:

correlation:

output format:  Send output to  Galaxy  GREAT

output file:  (leave blank to keep output in browser)

file type returned:  plain text  gzip compressed

To reset all user cart settings (including custom tracks), [click here](#).

**Intersect with SNPs (131)**

Select a group, track and table to intersect with:

group: Comparative Genomics track: Conservation  
 table: Primate Cons (phyloP4wayPrimates)

**Intersect SNPs (131) items with bases covered by Conservation:**

These combinations will maintain the names and gene/alignment structure (if any) of SNPs (131):

- All SNPs (131) records that have any overlap with Conservation
- All SNPs (131) records that have no overlap with Conservation
- All SNPs (131) records that have at least [80] % overlap with Conservation
- All SNPs (131) records that have at most [80] % overlap with Conservation

**Intersect bases covered by SNPs (131) and/or Conservation:**

These combinations will discard the names and gene/alignment structure (if any) of SNPs (131):

- Base-pair-wise intersection (AND) of SNPs (131) and Conservation
- Base-pair-wise union (OR) of SNPs (131) and Conservation

Check the following boxes to complement one or both tables. To complement a table means to:

- Complement SNPs (131) before base-pair-wise intersection/union
- Complement Conservation before base-pair-wise intersection/union

submit cancel

chr11	5201813	5201814	rs11036113	0	+
chr11	5201829	5201830	rs11036114	0	+
chr11	5201831	5201832	rs11036115	0	+
chr11	5201845	5201846	rs11036116	0	+
chr11	5201884	5201885	rs11036117	0	+
chr11	5201988	5201989	rs11036119	0	+
chr11	5201995	5201996	rs75541605	0	+
chr11	5201995	5201996	rs11036120	0	+
chr11	5202145	5202146	rs4910729	0	+
chr11	5202145	5202146	rs78065467	0	+
chr11	5202326	5202327	rs4910730	0	+
chr11	5202399	5202400	rs11036125	0	+
chr11	5202439	5202440	rs11036127	0	+
chr11	5202533	5202534	rs11036128	0	+
chr11	5202579	5202580	rs11036129	0	+
chr11	5202585	5202586	rs11036130	0	+
chr11	5202596	5202597	rs11036131	0	+
chr11	5202713	5202714	rs11036133	0	+
chr11	5202844	5202845	rs11036134	0	+
chr11	5202908	5202909	rs10837516	0	+
chr11	5202961	5202962	rs10837517	0	+
chr11	5203030	5203031	rs10837518	0	+
chr11	5203031	5203032	rs10837519	0	+
chr11	5203080	5203081	rs11036137	0	+
chr11	5203081	5203082	rs11036138	0	+
chr11	5203157	5203158	rs80276830	0	+
chr11	5203231	5203232	rs10837520	0	+
chr11	5203231	5203232	rs75845261	0	+

**To download upstream sequences for multiple human genes in a batch mode**

Ensembl

Home > Human (GRCh37) Location: X:37,200,583-43,606,068 Location-based display Chromosome X: 37,200,583-43,606,068

Dataset [None selected]

Dataset [None selected]

Ensembl Genes 59

- CHOOSE DATASET -
- CHOOSE DATASET -
- Danio rerio genes (Zv8)
- Gallus gallus genes (WASHUC2)
- Human sapiens genes (GRCh37)
- Mus musculus genes (NCBIM37)
- Rattus norvegicus genes (RGSC3.4)
- 
- Anolis carolinensis genes (AnoCar1.0)
- Bos taurus genes (Beta.4.0)
- Caenorhabditis elegans genes (WS210)
- Callithrix jacchus genes (caJacc3)
- Canis familiaris genes (CanFam.2.0)
- Cavia porcellus genes (cavPor3)
- Choloepus hoffmanni genes (choHof1)
- Ciona intestinalis genes (JGI2)
- Ciona savignyi genes (CSAV2.0)
- Dasyatis novemcinctus genes (dasNov2)
- Dipodomys ordii genes (dipOrd1)
- Drosophila melanogaster genes (BDGP5.13)
- Echinops telfairi genes (TENREC)



## To download upstream sequences for multiple human genes in a batch mode


The screenshot shows the Ensembl web interface. At the top, there are navigation links: Login, Register, BLAST/BLAT, BioMart, Tools, and More... Below this is a toolbar with buttons for New, Count, Results, URL, and XML. The main content area is divided into a left sidebar and a main panel. The sidebar has sections for Dataset (Homo sapiens genes (GRCh37)), Filters ([None selected]), Attributes (Ensembl Gene ID, Ensembl Transcript ID), and Dataset ([None Selected]). The main panel has a dropdown menu for 'Ensembl Genes by' set to 'Homo sapiens genes (GRCh37)'. Below this, there is a section titled 'Please restrict your query using criteria below' with expandable sections for REGION and GENE. The GENE section has a checkbox for 'Limit to genes...' with a dropdown menu set to 'with Illumina HumanWG 6 v1 ID(s)' and radio buttons for 'Only' and 'Excluded'. There is also a checkbox for 'ID list limit' with a text input field containing 'HGNC symbol(s) [e.g. ZFY]' and a 'Browse...' button. Two pink arrows point to the 'Count' button in the toolbar and the 'Filters' section in the sidebar.

The screenshot shows the Ensembl web interface with the 'Results' button highlighted in the toolbar. The main panel is titled 'Please select columns to be included in the output and hit "Results" when ready'. It features a list of checkboxes for various data types: Features, Structures, Transcript Event, Homologs, Variations, and Sequences. Under the 'SEQUENCES' section, there is a diagram of a gene structure and a list of checkboxes for 'Sequences (max 1)'. The 'Upstream flank' section has a checked checkbox for 'Upstream flank (1000)'. The 'Downstream flank' section has an unchecked checkbox for 'Downstream flank'. Below these are sections for 'Header Information', 'Gene Information', 'Transcript Information', and 'Exon Information', each with a list of checkboxes for specific data points. Two pink arrows point to the 'Count' and 'Results' buttons in the toolbar.

The screenshot displays the Ensembl genome browser interface. At the top, the Ensembl logo is visible along with navigation links for Login, Register, BLAST/BLAT, BioMart, Tools, and More... Below the navigation bar, there are tabs for New, Count, and Results. The main content area is divided into two columns. The left column contains a sidebar with sections for Dataset (Dataset 2 / 51737 Genes, Homo sapiens genes (GRCh37)), Filters (HGNC symbol(s), Upstream flank, Associated Gene Name), Attributes (Flank, Upstream flank, Associated Gene Name), and Dataset (None Selected). The right column shows an export interface with a dropdown menu set to 'File', a 'Go' button, and an 'Email notification to' field. Below this, there is a 'View' section with a dropdown set to '10' rows and a 'FASTA' format selection. The main content area displays a FASTA sequence for the HFE gene, starting with >HFE and followed by a long string of nucleotide characters.


## Step-by step instructions in the Genome Browsers class

## Obtain unique genes in a genome compared to other genome(s)



**JGI**  
DOE JOINT GENOME INSTITUTE  
OF ENERGY RESEARCH  
OFFICE OF SCIENCE

All Genomes  Quick Genome Search:  **GO**


INTEGRATED MICROBIAL GENOMES

IMG Home
Find Genomes
Find Genes
Find Functions
Compare Genomes
Analysis Carts
MyIMG
Using IMG

Gene Search
Cassette Search
BLAST
Phylogenetic Profilers

### Phylogenetic Profiler for Single Genes 3159 rows loaded.

Find genes in genome (bin) of interest qualified by similarity to sequences in other genomes (based on BLASTP alignments). Only user-selected genomes appear in the profiler.

Genome Completion: [F]inished, [P]ermanent Draft, [D]raft

#### Profile

Find Genes In'	With Homologs In	Without Homologs In	Ignoring	Taxon Name
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Archaea</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Crenarchaeota</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Aeropyrum</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Aeropyrum permix K1</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<b>Desulfurococcus</b>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Desulfurococcus kamchatkensis 1221n</a> [F]

<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<a href="#">Escherichia coli O157:H7 Sakai</a> [F]
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<a href="#">Escherichia coli O157:H7 str. TW14359</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli O157:H7 TW14588</a> [D]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli O26:H11 str. 11368</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli O55:H7 str. CB9616</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli S88</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli SE11</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli SE15</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli SMS-3-5</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli UMN026</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli UT189</a> [F]
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<a href="#">Escherichia coli W3110</a> [F]
<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<a href="#">Escherichia coli str. K-12 substr. MG1655</a> [F]

#### Phylogenetic Profiler for Single Genes Results 1022 genes retrieved

Processing 2 comparisons(s).  
 5373 genes found for genome (bin) of interest, Escherichia coli O157:H7 str. TW14359  
 1298 genes remaining after subtracting genes with homologs in Escherichia coli str. K-12 substr. MG1655  
 1022 genes remaining after intersecting with homologs in Escherichia coli O157:H7 Sakai


##### Summary Statistics

	Number	% of Total
Genes total number	1022	100.00%
COG	412	40.31%
Enzyme	23	2.25%
Pfam	678	66.34%
InterPro	308	30.14%
KO Term	359	35.24%
Tigfam	355	34.74%
No Functional Hit	205	20.04%
Unique In IMG	0	0.00%

Add Selected to Gene Cart
Select All
Clear All

Missing Gene? Tablath of the first selected gene in the list below against the genomes selected in Without Homologs in Genomes.

Select	Recall Flow	Gene Object ID	Locus Tag	Gene Name	Length	COG	Enzyme	Pfam	InterPro	KO Term	Tigfam	Unique In IMG
				putative type I fimbriae					IPR002259			



**img**


IMG Home | Find Genomes | Find Genes | Find Functions

Gene Search | Cassette Search | BLAST | Phylogenetic Profiles

### Genes assigned to COGs

**COG Categories**

- Amino acid transport and metabolism
- Carbohydrate transport and metabolism
- Cell cycle control, cell division, chromosome partitioning
- Cell motility
- Cell wall/membrane/envelope biogenesis
- Coenzyme transport and metabolism
- Defense mechanisms
- Energy production and conversion
- Extracellular structures
- Function unknown
- General function prediction only
- Inorganic ion transport and metabolism
- Intracellular trafficking, secretion, and vesicular transport
- Lipid transport and metabolism
- Posttranslational modification, protein turnover, chaperone
- Replication, recombination and repair
- Secondary metabolites biosynthesis, transport and catabolism
- Signal transduction mechanisms
- Transcription
- Translation, ribosomal structure and biogenesis



**img** INTEGRATED MICROBIAL GENOMES

IMG Home | Find Genomes | Find Genes | Find Functions | Compare Genomes | Analysis Cnts | MyIMG | Using IMG

Gene Search | Cassette Search | BLAST | Phylogenetic Profiles

### Genes assigned to COG: Intracellular trafficking, secretion, and vesicular transport

Add Selected to Gene Cart | Select All | Clear All

Search column: Gene ID | Search term:

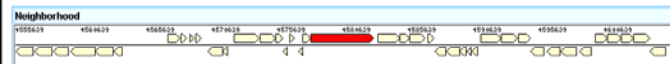
Export | Page 1 of 1 | << first < prev 1 next > last >> | All

Column Selector | Select Page | Deselect Page

Select	Gene ID	Gene Name	Genome
<input type="checkbox"/>	644919529	putative type-1 fimbrial protein	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644919643	putative fimbrial protein	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644919870	putative beta-barrel outer membrane protein	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644920051	putative outer membrane export protein	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644920058	hypothetical protein	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644920341	putative protease encoded in prophage CP-932K	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644920509	capsid assembly protein	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644920596	putative outer membrane transporter of ShigaHecA/Pha exoproduct family	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644920790	hemolysin activator protein precursor	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644921019	putative capsid protein of prophage CP-932K	Escherichia coli O157:H7 str. TW14359
<input type="checkbox"/>	644921162	protein C (EC:3.4.21.69). Serine peptidase. MEROPS family S49 (M0fam)	Escherichia coli O157:H7 str. TW14359

### Evidence For Function Prediction

**Neighborhood**



red = Current Gene  
green = Positional Cluster Gene in the same KEGG Pathway as the Current Gene  
||||| CRISPR array

[Sequence Viewer For Alternate ORF Search](#)

Chromosome Viewer colored by:  Select Function

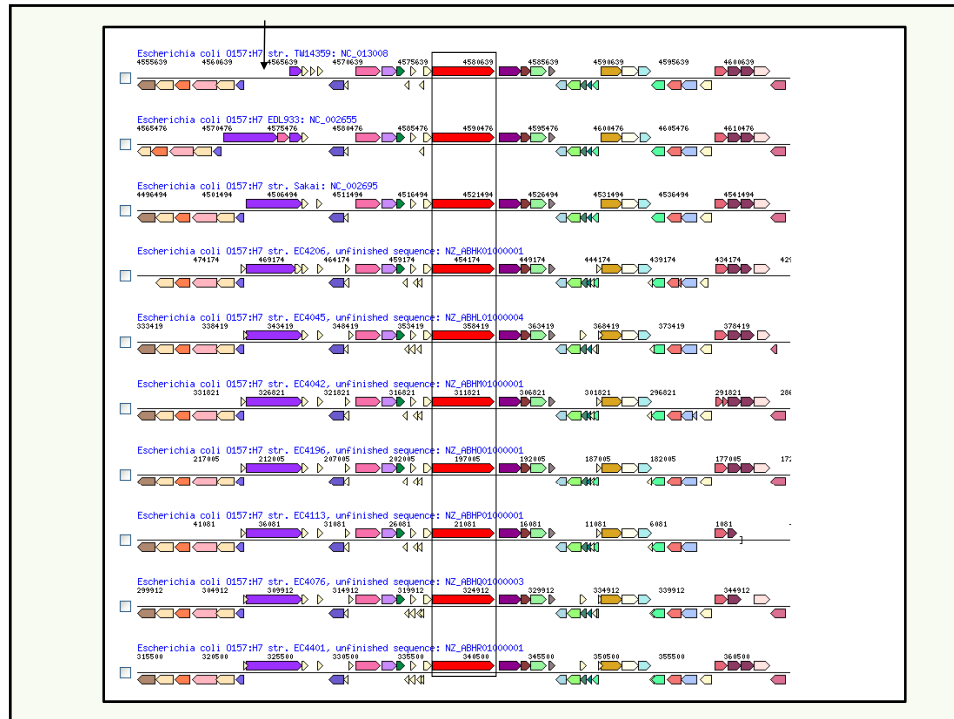
**Conserved Neighborhood**

[Show ortholog neighborhood regions](#)

Chromosomal Cassette Viewer By:  Select Protein Cluster

COG ID	Consensus Sequence Length	Description	Percent Identity	Alignment On Query Gene	E-value	Bit Score
COG5295	715	[M] Extracellular structures [M] Intracellular trafficking, secretion, and vesicular transport Autotransporter adhesin	34.8		2.0e-35	146

Pfam Domain	HMM Pfam Hit	Description	Percent Alignment On Query Gene	Alignment On Query Gene	E-value	HMM Score
Hep_Hag	pfam05658	Hep_Hag	1.76		2.7e-04	20
Hep_Hag	pfam05658	Hep_Hag	1.76		1.4e-05	24
Hep_Hag	pfam05658	Hep_Hag	1.76		2.3e-05	24
Hep_Hag	pfam05658	Hep_Hag	1.70		1.7e-05	24
Hep_Hag	pfam05658	Hep_Hag	1.76		1.2e-06	28
Hep_Hag	pfam05658	Hep_Hag	1.76		4.0e-06	26
Hep_Hag	pfam05658	Hep_Hag	1.76		9.8e-06	25
Hep_Hag	pfam05658	Hep_Hag	1.76		5.2e-05	22
HIM	pfam05662	Haemagglutinin	1.51		4.1e-06	26
HIM	pfam05662	Haemagglutinin	1.51		1.7e-07	30
HIM	pfam05662	Haemagglutinin	1.51		8.2e-08	31
HIM	pfam05662	Haemagglutinin	1.51		7.2e-08	31
Hep_Hag	pfam05658	Hep_Hag	1.76		8.0e-07	28



## Practical examples

- Non-synonymous SNPs that are associated with diseases and with known 3-D structures of the proteins (NCBI)
- SNPs in the conserved regions (UCSC)
- Download upstream sequences for multiple human genes (Ensembl)
- Obtain unique genes in a genome compared to other genome(s) (IMG)

## Training

Making Sense of DNA and Protein Sequences

Gene Resources: From Transcription Factor  
Binding Sites to Function

Sequence Similarity Search: BLAST

Protein Structural Analysis: Binding Sites to  
Distant Homologs

Genome Browsers

Identification of Disease Genes

Correlation of Disease Genes to Phenotypes

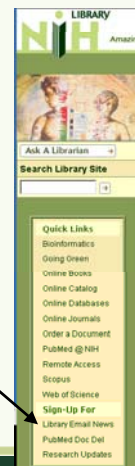
Currently offering as an FAES course

“Practical Bioinformatics”

## Training

September 30	GeneSpring 11.0.2
October 27-28	Partek Genomics Suite
November 2	Lasergene and SeqMan Ngen
	ArrayStar/Qseq (DNASTAR)
November 9-10	Genomatix

<http://nihlibrary.nih.gov/bioinformatics>



## One-on-one consultations

Wide range of question  
Identifying a correct resource  
Collaboration on a research project

Gene set enrichment analysis from  
microarray experiments  
Download upstream gene sequence  
and identify transcription factor binding  
sites



Bioinformatics specialist with  
programming expertise  
Dr. Lynn Young



Microarray data analysis  
Next gen sequence analysis

## Outline

- Databases and Tools
- Licensed resources from the NIH Library
  - Dedicated computers
- Practical examples
- Training
- One-on-one consults
  - Programming help



Suggestions/Questions/Comments

Medha Bhagwat  
bhagwat@mail.nih.gov





## Opening New Doors to Information



LIBRARY  
Amazing Research. Amazing Help.

Home | Library Services | Research Tools | Resource Training | About Us

For the Public | FAQ | Help | Site Map

HBRL Users Click Here

### Welcome to The NIH Library

An ODS Service

Ask A Librarian

Search Library Site

All Research Tools A-Z

ABCDEFGHIJKLMNOPQRSTUVWXYZA

Quick Links

- Bioinformatics
- Going Green
- Online Books
- Online Catalog
- Online Databases
- Online Journals
- Order a Document
- PubMed @ NIH
- Remote Access
- Scopus
- Web of Science
- Sign-Up For
- Library Email News
- PubMed Doc Del
- Research Updates

Popular Databases

- PubMed
- Web of Science
- Scopus
- UpToDate
- Micromedex
- View all Databases

Online Journals

- Nature
- Science
- JBC
- JAMA
- NEJM
- View all Journals

Librarian Picks

- Bioinformatics
- Writing Center
- Renew Books
- Online Catalog
- Online Books
- View all Services

FIRST STEP Search for articles, news, videos and images by category

NIH Library News Features

**An Introduction to Genomics Resources and their Practical Applications Seminar September 28, 12-1PM**

Dr. Medha Bhagwat, NIH Library Bioinformatics Program Coordinator, will present a seminar on "An Introduction to Genomics Resources and their Practical Applications" on September 28 at noon. Due to an overwhelming response to a similar presentation arranged by the NIH Library in July, this seminar will be offered in the larger Lipsett auditorium. Attendees of that previous presentation will also learn some additional information and practical tips.

The seminar will provide:

**PATAI'S Chemistry of Functional Groups Available Online**

The most comprehensive work in functional groups is now available as a fully searchable online resource with additional functionality and search capabilities. Patai covers all aspects of the chemistry of functional groups – an essential tool for the organic chemist – including physical organic chemistry, analytical chemistry and techniques, reaction mechanisms, chapters on synthetic pathways, reactions and strategies as well as applications in drug discovery, pharmaceuticals, biochemistry and molecular biology.

Read More

Announcements

- NIH Library Classes: August and September 2010
- Launch of Beta Search Engine: First Step
- Locked Carrels 2010
- NIH Library All Announcements
- View All Announcements

Upcoming Training Classes

- Discover Scopus and Web of Science: Complementary Giants Sep 23
- Web Search: Thinking Beyond Google Sep 23
- PubMed: Understanding the Basics Oct 13

Request a Tutorial

http://nihlibrary.nih.gov/

<http://nihlibrary.nih.gov/bioinformatics>

**Thanks!**

Medha Bhagwat  
bhagwat@mail.nih.gov

